Allowing camera tilts for document navigation in the standard GUI: a discussion and an experiment

Yves Guiard, Olivier Chapuis, Yangzhou Du, Michel Beaudouin-Lafon

Allowing Camera Tilts for Document Navigation in the Standard GUI: A Discussion and an Experiment

Yves Guiard [a] Olivier Chapuis [b] Yangzhou Du [b] Michel Beaudouin-Lafon [b]

[a] Laboratoire Mouvement et Perception
CNRS & Université de la Méditerranée, Marseille, France
yves.guiard@univmed.fr

[b] LRI & INRIA Futurs
Université Paris-Sud, Orsay, France
{chapuis, du, mbl}@lri.fr

ABSTRACT
The current GUI is like a flight simulator whose camera points fixedly at right angle to the document, thus preventing users from looking ahead while navigating. We argue that perspective viewing of usual planar documents can help navigation. We analyze the scale implosion problem that arises with tilted cameras and we report the data of a formal experiment on document navigation with perspective views.

Categories and Subject Descriptors
H.5.2. [**User Interfaces**]: Interaction styles; I.3.6. [**Methodology and Techniques**]: Interaction techniques.

Keywords
Perspective viewing, camera tilt, multiscale document navigation, Fitts' law.

1. INTRODUCTION

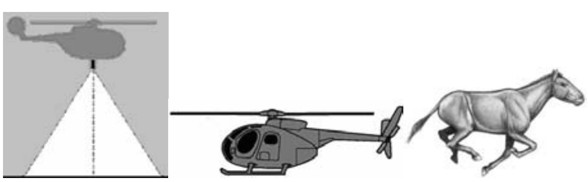

Fig. 1. Perpendicular viewing with a vertically oriented camera (A) vs. forward vision in real aircrafts (B) and animals (C).

Imagine that the owner of a model helicopter has just purchased a wireless miniature video camera to enjoy the flight as if onboard. How should the camera be mounted on the aircraft? Certainly not with a downward, vertical orientation, as shown in Fig. 1A. During self-motion one needs to see ahead. Wind screens (Fig. 1B) as well as the visual equipment of animals (Fig. 1C) *always* point forward, and so do virtual cameras in flight simulators.

Yet, computer users are forced to navigate all sorts of documents with a virtual camera whose orientation remains fixedly perpendicular to the document surface, as shown in Fig. 1A. As far as document navigation is concerned (i.e., moving to another region of the document as distinct from working locally), a graphical user interface (GUI) is like a flight simulator. Navigating a document with the current technology means moving a virtual camera [7] mounted on an implicit virtual—i.e., immaterial hence gravity-free, yet remote-piloted—aircraft. Zooming in or out and scrolling mean translating the virtual camera-aircraft system upward or downward and in a plane parallel to the document, respectively. It seems important to realize that our standard GUI, which simulates a remote-piloted aircraft with a closed-circuit video system, is an instance of a (non-immersive) virtual-reality situation—but one that suffers its own original sin, a fixed upright orientation of the virtual camera.

2. PREVIOUS WORK
Of course there is room for facilitating navigation with perpendicular views of planar documents [2,12], but let us ask about existing solutions to the design problem depicted in Fig. 1A. Rather than considering the camera orientation issue, HCI research has focused on a revision of the spatial arrangement of the documents, moving from the familiar planar surface to a diversity of 3D-space layouts. Authors have mainly investigated the environmental metaphor of the office [3,13,14,15], the object metaphor of the book [4], or combinations [5]. Even though some recent map navigators do allow perspective viewing with freely tiltable cameras (e.g., Google Earth), the HCI literature does not appear to have inquired so far into the possibility of camera tilts to help users navigate all their familiar planar documents (text pages, spreadsheets, pages of code, web pages, etc.).

3. NAVIGATING WITH A TILTED VS. PERPENDICULAR CAMERA
3.1. Varieties of Camera Rotations
Assuming the user approximately knows in which direction to search for some remote object, one obvious technique is to tilt the camera until the area of interest enters the view—the *panoramic* rotation of film makers (Fig. 2A), but there are many other possible camera tilt techniques. For example, if the user wants to peek at a distant region without losing sight of the local work, one attractive variant is what we call the *lunar* rotation[1]—revolving the aircraft along a half circle

[1] The *lunar* metaphor refers to the property that the moon always faces the Earth from the same angle, revolving about our planet at the same pace as it rotates about itself.

with the camera being constraining to remain oriented towards its current fixation point in the document (Fig. 2B). Another useful rotation is what we call the *trans-rotation*—the camera translates at a fixed altitude while keeping pointed to a constant fixation point (Fig. 2C).

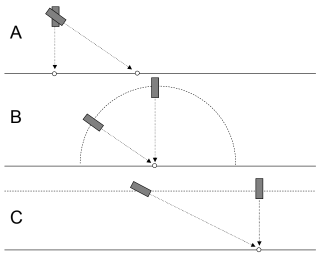

Fig. 2. Three kinds of camera rotation: the panoramic rotation (A), the lunar rotation (B), and the trans-rotation (C).

3.2. Screen View vs. Document Selection

Let us distinguish two ways in which the mapping of document space to view space can be thought of (Fig. 3). While in screen space we can describe the view displayed to the user (Fig. 3AB), it is only in document space (Fig. 3CD) that the full mapping can be described, showing not only what we call the *selection* (white area) but also the subset of the document that is not visualized (dark area).

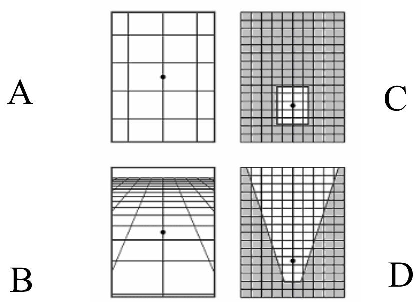

Fig. 3. Effect of a lunar rotation of the camera as described in screen space (from A to B) and in document space (from C to D).

3.3. Geometric Advantages of Perspective Viewing

Perspective viewing (PV) is inherently multi-scale. The visualization scale varies continuously within the view, making it possible to simultaneously display the local detail and the remote context [13]—obviously an attractive feature when it comes to very large documents. Unlike the zooming technique, based on a time variation of scale, PV allows a whole range of visualization scales to be available at once. Compared with other techniques based on spatial multiplexing such as the bifocal view [1] and the fisheye [8], PV offers several advantages. The scale variation is gradual, rather than abrupt, allowing a reliable, faultless mapping of the document selection to the view. Second, PV rests on a visualization scale variation with distance that is simple, monotonic, and non-arbitrary, being based on the natural laws of ecological optics [9]. Third, the range of scales that can be represented with PV is very large, allowing this technique to adapt to arbitrarily large documents, as is the case with the usual pan and zoom (P&Z) technique [10]. Finally, PV is familiar, since it is in perspective that we see most surfaces of the real world.

3.4. Looking Out for Information: Saving Screen Pixels

Suppose you are editing a line of text (the dotted line in Fig. 4) in a linearly arranged document (the black rectangle) and you realize you need to check some heading in the beginning of the document. What you currently see in document space is selection S_1 (small gray unfilled rectangle) but what is now of interest to you is the out-of-view area marked with a lighter shade (degree of interest along the document is coded with shades of gray). The state-of-the-art solution (Fig. 4A) is to zoom out until the region of interest is included. But scaling up the selection from S_1 to S_2 means expanding it indistinctly in all directions, irrespective of the user's intention—recall that you actually wanted to look in the *upward* direction. Most of S_2 is irrelevant (including a lot of background void plus all dark, uninteresting regions of the document), meaning a dramatic waste of screen pixels. Second, the particular line you were just editing (the dotted line) is no longer visible because the viewing scale has dropped uniformly everywhere. In contrast, Fig. 4B shows an upward tilt of the camera (a lunar rotation), which avoids the above shortcomings. The initial rectangle S_1 turning into trapezoid S_3, note that the obtained selection includes very little background void and excludes most of the unattended material; second, the detail of the previous selection can still be read and edited because for that region the scale level has been conserved (an interesting property of the lunar rotation).

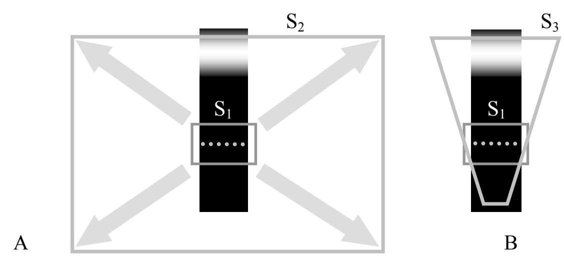

Fig. 4. Selection change with a zoom out (A) vs. a lunar rotation (B).

3.5. Navigating to One's Target: Providing the Pilot with Prospective Information

Having managed to see your target, with either a zoom out or a camera tilt, your task now is *to go there* using the visual information from your perpendicular or tilted camera.

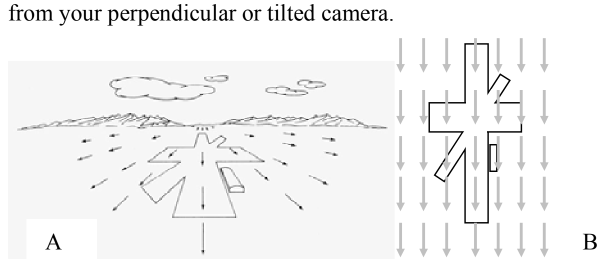

Fig. 5. The optical flow field available to someone onboard an aircraft when looking ahead (A) and downward (B). Fig. 5A reproduced from J. J. Gibson [9], Fig. 7.4 p. 124.

Fig. 5A shows the structure of the optical flow field available from onboard an aircraft when looking ahead. The radial optical expansion, whose focus coincides with the horizon, tells you that the aircraft is going to cross, at the same altitude, the gap between the two faraway hills [9]—thus while looking ahead one obtains the *prospective* visual information one needs to fly the aircraft. Fig. 5B shows the laminar structure of the optical flow field that would be available to someone (hopefully not the pilot) looking to the ground: one has no sense of where one is going. Note that this is just the kind of pattern experienced in GUIs during scrolling. Even though electronic document navigation is safer than real flying, there is reason to question the suitability of a fixedly upright camera, keeping in mind that the user is the pilot of the GUI.

4. THE MAIN DIFFICULTY OF PV: SCALE IMPLOSION

4.1. The Virtual Camera Model

Fig. 6 illustrates—in 2D space for simplicity—the virtual camera model with the camera tilted at angle α. The observation point is O and the field of view is the white cone AOC. The horizontal solid line at the bottom represents the document surface. The tilted solid line stands for the projection plane (the screen). The selection is shown as segment AC. The view—the screen subset dedicated to the visualization—is shown as segment A'C'. If view size is a constant, then viewing angle AOC is determined by the camera's focal length h = OI. The cross at B' stands for the screen cursor (constrained to stay within interval A'C'), which serves to specify or grasp a point (B) within the document selection. The figure also illustrates a target (segment DE)—some currently invisible text element or graphical object that the user wants to reach through navigation. When distance O'D tends to infinity, the *visualization scale*, defined as the D'E'/ DE ratio, tends to zero.

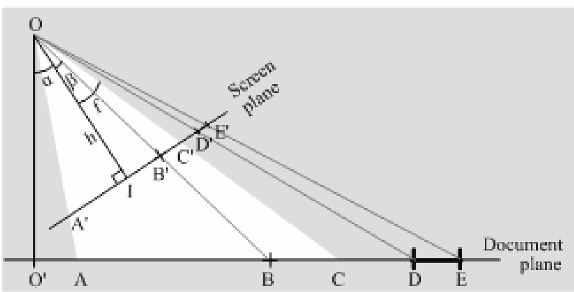

Fig. 6. The virtual camera model, with both translations and rotations allowed, reduced to 2D space.

4.2. Scale Implosion in PV

Defining the view by its half-size v = IC' and its half-angle f, we have $\tan f = v/h$. Without loss of generality, we assume that OO'=h. If α≠0, a point at coordinate IB' = x in the view corresponds to a point at coordinate O'B = $p(x)$ in the document. Using trigonometric calculations we get

$$p(x) = \frac{x + h\tan\alpha}{1 - \frac{x}{h}\tan\alpha} \qquad (1)$$

Deriving $p(x)$ gives the inverse of the local scale $S(x)$ of the document at the point of coordinate x in the view:

$$S(x) = \frac{dp(x)}{dx} = 1/\left(\cos\alpha - \frac{x}{h}\sin\alpha\right)^2 \qquad (2)$$

Plotting $S(x)$ shows the scale implosion as the x coordinate gets closer to the horizon (Fig. 7A). Scale is close to 1 in the bottom half of the view ($x \in [-v, 0]$). This means that a click-and-drag technique like that of Adobe Acrobat Reader may be suitable for navigation: click on the faraway target, then drag it home until the corresponding section of the document reaches the bottom of the view where it has a scale of about one. The problem is that close to the horizon screen pixels represent huge document areas. If the area represented by one screen pixel is larger than half the view size, when the user clicks that pixel and drags down, the corresponding section of the document is enlarged. If the section gets larger than half the view, the target may get out of the view or may be still too small and the user may miss it, requiring additional click-and-drag actions to reposition the target. This situation occurs when the scale at the top of the view $S(v)$ is larger than half the view size v. We now evaluate the minimum index of difficulty at which this occurs, i.e. such that $S(v) = v / 1 = v$.

For a position d in the document the minimum rotation angle α needed to bring that portion of the document into the view is given by:

$$\alpha = \arctan\left(\frac{d - v}{h + dv/h}\right) \qquad (3)$$

Using Equations (2) and (3), we can compute the distance d_{max} such that $S(v) = v$ and therefore, assuming targets of minimal size 8 (as in our experiment), the index of difficulty ID_{max} beyond which targets cannot be reliably selected in a single click-and-drag action. For a typical display (v = 512, f = 30º, h = 295), we get α = 57.08º, d = 23169 and ID_{max} = 11.5 (this theoretical limit was confirmed in our experimental data: The average number of drags for IDs 9, 11, 13 and 15 were respectively 1.04, 1.11, 1.81 and 2.59 for PV and 1.17, 1.27, 1.64 and 2.28 for P&Z).

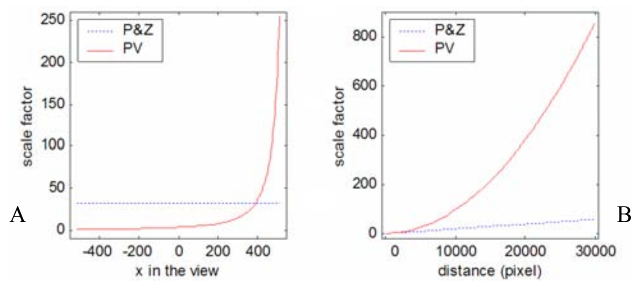

Fig. 7 - Scale implosion for PV (note that the vertical axis shows the inverse of the visualization scale).

It is also enlightening to compare the visualization scale of the target when it enters the view for the PV vs. P&Z techniques. Fig. 7B shows the inverse of the scale factor at the target against target distance. Scale implodes for PV, but the growth is linear for P&Z: so PV is likely to work less and less well relative to P&Z as the target gets farther and farther away.

5. AN EXPERIMENT ON DOCUMENT NAVIGATION WITH PV

Because of the scale implosion inherent in PV, target reaching with a tilted camera should *not* obey Fitts' law—one expects a concave up curvature in the MT vs. ID function. Our experiment was designed primarily to test this hypothesis while providing some sense of how PV performance compares with the familiar P&Z technique.

5.1. Methods

We used Fitts' pointing task [10]. Fitts' law links target-acquisition time (MT) to the ratio of target distance (D) to target width (W), namely, $MT = k_1 + k_2 * \log_2(D/W + 1)$, where k_1 and k_2 are adjustable coefficients (k_2>0) and $\log_2(D/W + 1)$ stands for the task *index of difficulty* (ID). With D and W measured in document space (O'D and DE in Fig. 6), the ID measures task difficulty independently of the navigation technique used to zero out distance O'D.

We used a standard wheel mouse and the familiar click-and-drag scrolling technique. The experiment was run in full-screen mode on a 19' monitor. The targets were two constant-size (8-pixels high, 283 pixels wide) blue-filled rectangles drawn at varying distances one above the other on an oblong rectangular document (height to width ratio = 1,048,560 to 104,856 pixels). The document was covered with an array of black concentric circles centered around the target. At least one arc was visible from any position at the highest scale, hence disorientation was precluded. Clicking the target caused the concentric pattern to be instantly rearranged around the other target. Whenever the target was less than 3 pixels high in the view, the program replaced the zoomable blue target with a green 2-pixel thick horizontal line (the 'beacon').

The target being initially located out of view, the participants' task was to first visualize the distant target (shown as a beacon) with the mouse wheel, by either zooming-out or tilting the camera (in both cases one wheel notch enlarged/reduced the selection by 10%) and then to navigate to the target with the click-and-drag technique.

Both techniques were tested with four *ID*s: 9, 11, 13 and 15bits (i.e., up to a *D*/*W* ratio of 32,767). Participants were instructed to perform as fast as possible but any click error had to be corrected at once. Hence error rate was a constant 0% in our data and *MT* was defined as the time elapsed between two consecutive successful target clicks. Sixteen adult volunteers participated in two 40-mn sessions, one for each technique. Ignoring warm-up trials, there were 24 actual measurements of *MT* per level of *ID* and per participant for each technique.

5.2. Results and Discussion

For the P&Z condition, the *MT* vs. *ID* relationship was essentially linear (Fig. 8), with all individual r^2 values in the .915-.999 range, in keeping with previous findings [10]. In contrast, the PV curve exhibits a distinctive upward concavity, well fitted by an exponential, present in all 16 subjects. This expected curvature confirms that using this bare, unaided implementation of PV, navigation is likely to fail for much higher levels of *ID* due to unacceptably long *MT*s. Note, however, that our Fitts' law test covered a rather large range of *ID*s. While moving from the first to the last verse of Shakespeare's complete works (i.e., 150,000 verses) arranged as a single linear document would involve an *ID* of 17.2 bits [11], most of the documents we actually use everyday are far smaller than that.

We ran a two-way repeated measures ANOVA on log(*MT*) with the technique (P&Z vs. PV) and the *ID* (9, 11, 13, 15 bits) as factors. Beside the trivial effect of the *ID* ($F_{3,45}$=840, p>.0001), we found a significant interaction ($F_{3,45}$=29, p>.0001), but no main effect of the technique (F<1). For *ID* = 15bits, performance was better with the P&Z technique (t_{15} = 5.77, p = .005, Bonferroni correction), but for *ID* = 9bits the reverse was true (t_{15} = 3.88, p = .006).

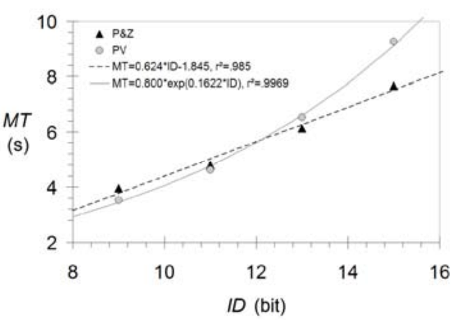

Fig. 8. *MT* vs. task difficulty for the two conditions.

6. IMPLICATIONS FOR FUTURE RESEARCH

This contribution mainly aims at calling the attention of the HCI research community to the potential utility of one or two camera tilt degree(s) of freedom for navigating, not just geographical maps, but the whole diversity of planar documents we handle everyday in usual GUIs. With the progress of the hardware technology (notably, graphics cards and OpenGL programming), camera tilts are no longer difficult to implement today. The real problem ahead is one of designing system-aided PV implementations that will be, if possible, more efficient than the perpendicular viewing techniques of the state of the art [6].

Yet, to obtain a full evaluation of the PV technique for document navigation, the multiscale pointing paradigm [10] is unlikely to suffice. Beside the issue of target-reaching efficiency, one noteworthy advantage of PV that cannot be captured in a target-acquisition performance score is that during the navigation process—which may last quite a few seconds for difficult reaching tasks—the user is in position to receive far richer information from the document than with the state of the art techniques. For example, using the P&Z technique, most of the document can be overviewed, but not really seen. In contrast, traveling across a document with PV means visiting or exploring information space, which is more than just traversing it. Thus one plausible hypothesis for future research is that during long-distance navigation users learn more with the PV than P&Z technique about the document they traverse. To test this, a new experimental paradigm probably needs to be designed to quantify the acquisition of knowledge about the contents of the document during goal-directed navigation.

7. ACKOWLEGMENTS

We gratefully acknowledge the generous financial support of the French *Ministère délégué à la recherche et aux nouvelles technologies* (*ACI Masses de données*, Micromegas Project).

8. REFERENCES

1. Apperley, M. & Spence, I.T.R. 1982. A bifocal display technique for data presentation. *Proc. Eurographics*, 27-43.
2. Appert, C. & Fekete J.-D. In press. OrthoZoom Scroller: 1D Multi-Scale Navigation. *Proc. of CHI 2006, ACM.*
3. Ballay, J.M. 1994. Designing Workscape: An interdisciplinary experience. *Proc. CHI'94,* ACM, 10-15.
4. Card, S.K., Hong, L., Mackinlay, J.D., & Chi, E.H. 2004. 3Book: A scalable 3D virtual book. *Ext. Abstracts CHI '04,* ACM 1095-1098.
5. Card, S.K., Robertson, G.G., & York, W. 1996. The WebBook and the Web Forager: an information workspace for the World-Wide Web. *Proc. CHI'96*, ACM, 111-ff.
6. Du, Y., Guiard, Y., Chapuis, O., & Beaudouin-Lafon, M. In press. Assisting Target Acquisition in Perspective View. *Proc. British HCI'06*. London, (UK), Sept. 2006.
7. Foley, J.D., van Dam, A., Feiner, S.K., & Hughes, J.F. 1990 (2dn Ed.). *Computer graphics: Principles and practice*. Boston, MA, Addison-Wesley Longman.
8. Furnas, G.W. 1986. Generalized fisheye views. *Proc. CHI '86,* ACM, 16-23.
9. Gibson, J.J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
10. Guiard, Y. & Beaudouin-Lafon, M. 2004. Target acquisition in multiscale electronic worlds. *Internat. J. Hum.-Comp. Studies 61*(6), 875-905.
11. Guiard, Y., Du, Y., Fekete, J.-D., Beaudouin-Lafon, M., Appert, C., & Chapuis, O. In press. Shakespeare's complete works as a benchmark for evaluating multiscale document-navigation techniques. *Proc. BELIV06, a Workshop of the AVI 2006 International Working Conference.*
12. Igarashi, T., Hinckley, K. 2000. Speed-dependent automatic zooming for browsing large documents, *Proc. UIST'00*, 139-148.
13. Mackinlay, J.D., Robertson, G.G., & Card, S.K. 1991. The perspective wall: detail and context smoothly integrated. *Proc. CHI'91.* ACM, 173-176.
14. Robertson, G., van Dantzich, M., Robbins, D., Czerwinski, M., Hinckley, K., Risden, K., Thiel, D., & Gorokhovsky, V. 2000. The Task Gallery: a 3D window manager. *Proc. CHI '00*. ACM, 494-501.
15. Robertson, G.G. and Mackinlay, J.D. 1993. The document lens. *Proc. UIST '93*. ACM, 101-108.